

## RESEARCH ARTICLE

# Shotgun identification of proteins from uredospores of the bean rust *Uromyces appendiculatus*

Bret Cooper<sup>1</sup>, Wesley M. Garrett<sup>2</sup> and Kimberly B. Campbell<sup>1</sup>

<sup>1</sup> Soybean Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, USA

<sup>2</sup> Biotechnology and Germplasm Laboratory, USDA-ARS, Beltsville, MD, USA

We are interested in learning more about the proteome of *Uromyces appendiculatus*, the fungus that causes common bean rust. Knowledge of the proteins that differentiate life-cycle stages and distinguish infectious bodies such as uredospores, germlings, appressoria, and haustoria may be used to define host–pathogen interactions or serve as targets for chemical inhibition of the fungus. We have used 2-D nanoflowLC-MS/MS to identify more than 400 proteins from asexual uredospores. A majority of the proteins appear to have roles in protein folding or protein catabolism. We present a model by which an abundance of heat shock proteins and translation elongation factors may enhance a spore's ability to survive environmental stresses and rapidly initiate protein production upon germination.

Received: August 29, 2005  
Revised: October 16, 2005  
Accepted: November 6, 2005

**Keywords:**

MudPIT / Proteome / Rust / Spore / *Uromyces appendiculatus*

## 1 Introduction

Rust fungi of the order Uredinales are obligate pathogens that infect plants such as wheat and beans [1]. Rusts have complex life cycles and heteroecious rusts complete their sexual stages through infection of secondary hosts. Rusts can also asexually propagate by producing uredospores from uredia, the fruiting bodies that appear as pustules on infected leaves. Uredospores are wind-borne and are resistant to thermal extremes. In the presence of moisture and 18–22°C temperature, uredospores germinate and spore germ tubes grow across the leaf surface until a stomata [1, 2] or, in some cases, a suitable epidermal cell is encountered [3]. An appressorium then forms that allows the fungus to penetrate the stomata opening or through the epidermis directly and the fungus begins to colonize the interior of the leaf and adjacent cells, ultimately leading to the formation of uredia

that break through the leaf surface and release more uredospores [1, 2]. Rapid spread, parasitism, and destruction of the host make rust fungi formidable pathogens.

Rust fungi are notoriously difficult to study because they are obligate pathogens. Uredospores will germinate in a drop of water and appressoria formation can be induced on polystyrene leaf-surface replicas [4], but the formation of other infectious structures such as haustoria, the organ that absorbs nutrients from a cell, or uredia have not been readily attained *in vitro*. Despite the discovery of a few genes involved in pathogenesis and nutrient uptake [5–8] and documentation of transient transformation and gene expression [9, 10], the intractability of these organisms has stalled further understanding of the molecular biology of rusts. Similar limitations had once slowed research advancements on the partially obligate malaria parasite, until its genome was sequenced and major proteomics efforts were undertaken to identify proteins associated with different steps in its life cycle [11]. As a result, malaria research is now moving at an accelerated pace [12].

We have begun to decipher the proteome of *Uromyces appendiculatus*, the causal agent of bean rust, and are starting with the uredospore. The knowledge of the proteins that form uredospores may be used to understand spore survivability or be used in the design of chemical inhibitors of pro-

**Correspondence:** Dr. Bret Cooper, USDA-ARS, 10300 Baltimore Ave., Bldg. 006/Rm. 213, Beltsville, MD 20705, USA

**E-mail:** cooperb@ba.ars.usda.gov

**Fax:** +1-301-504-5728

**Abbreviations:** MudPIT, multidimensional protein identification technology; NR, nonredundant

teins active in spore germination. We produced MS/MS spectra using 2-D nanoflow LC-MS/MS, also known as multidimensional-protein identification technology (MudPIT) [13], and used MASCOT [14] to identify peptides from protein sequences in the NCBI nonredundant (NR) protein database. Coupled with DBParser [15], a program that parsed the assembled protein data into a NR dataset, we have defined a probability-based and parsimonious sum of inferred proteins for *U. appendiculatus* uredospores.

## 2 Materials and methods

### 2.1 Purification of proteins

A liquid suspension of *U. appendiculatus* race 41 uredospores was sprayed onto *Phaseolus vulgaris* cv. Early Gallatin plants using a compressed-air aerosol-spray canister. Plants were placed in a dew chamber overnight (18°C) and then grown in a growth chamber (26°C/90% relative humidity/12 h days) until uredospores emerged from pustules on the leaves. Spores were collected with a Cyclone Surface Sampler (Burkard Manufacturing, Rickmansworth, UK) and then filtered through a 106 µm mesh screen. Isolated uredospores were then frozen in liquid nitrogen and pulverized with a mortar and pestle. Proteins were precipitated in acetone/TCA [16], resolubilized in 8 M urea/100 mM Tris–HCl pH 8.5, reduced in trichloroethylphosphine and carboxyamidomethylated in iodoacetamide [17]. After digesting 1 mg soluble protein with endoproteinase Lys-C at 37°C for 12 h (Roche Applied Science, Indianapolis, IN), the reaction was diluted to 2 M urea with 100 mM Tris–HCl pH 8.5, adjusted to 2 mM CaCl<sub>2</sub> and digested with Porozyme immobilized trypsin (Applied Biosystems, Foster City, CA) at 37°C for 12 h as previously described [13]. Immobilized trypsin was removed by centrifugation. All peptides were separated from spore debris using 0.45 µm PVDF filters (Millipore, Bedford, MA) and were concentrated by SPE using SPEC-PLUS PT C18 columns (Varian, Lake Forrest, CA) followed by centrifugal vacuum evaporation.

### 2.2 LC separation

Columns were prepared from 365 od × 75 id fused-silica capillaries (Polymicro Technologies, Phoenix, AZ). Each capillary was drawn to a 5 µm tip using a P-2000 laser puller (Sutter Instrument, Novato, CA) [18]. Then, 9 cm of 5 µm Aqua reverse-phase C18 resin (Phenomenex, Torrance, CA) followed by 4 cm of 5 µm Luna strong cation exchange resin (Phenomenex) were packed under 600 psi using a helium pressure cell [19]. The peptides were also loaded onto the packed column using the same pressure cell. The loaded column was then placed in-line with a Surveyor HPLC pump that is part of the ProteomeX workstation (Thermo Electron, Waltham, MA). Peptides were eluted in a 12-step process that included increasing concentrations of salt followed by an

increasing gradient of mobile phase at each step, as previously described [20]. Flow rate was 200 nL/min. The eluent was electrosprayed directly into the ESI source of the LCQ-Deca XP IT mass spectrometer. Electrospray voltage (1.8 kV) was applied at a liquid junction before the column *via* a gold electrode [21].

### 2.3 MS/MS

A parent-ion scan was performed over the range 400–1600 *m/z*. Automated-peak recognition, dynamic exclusion, and MS/MS-ion scanning of the top three most intense parent ions were performed using the Xcalibur 1.3 software as described previously [22, 23]. The tandem mass spectra were extracted from the raw data by Bioworks 3.1. Parameters were set at: 400 minimum mass, 3500 maximum mass, 15 minimum peaks, 100 000 minimum TIC, 1.4 Da precursor mass tolerance, 25 intermediate scans, and one group scan. All spectra not calculated as being singly charged were extracted as both doubly and triply charged spectra. A merge.pl script that is part of the MASCOT 2.0 software package (Matrix Science, London, UK) was used to convert multiple .dat files into a single file suitable for searching.

### 2.4 Peptide sequence inferences from spectra

54 535 MS/MS spectra from uredospores were evaluated by MASCOT 2.0 [14] and compared to virtual spectra created from peptides derived from the 01/19/2005 release of the NCBI-NR protein database (2 308 679 sequences; ftp://ftp.ncbi.nih.gov/blast/db/). Searches were performed on a seven-node 3.2 GHz Dell server. Search parameters were set at fully tryptic digests and zero missed cleavages; carboxyamidomethylation was selected as a fixed mass modification; oxidation (M), N-term pyro-cmC, and N-term pyro-Glu (E,Q) were set as variable-mass modifications. Mass values were averaged, and peptide mass tolerance and fragment mass tolerance were set at ±1.5 Da and ±0.8 Da, respectively.

### 2.5 Assembly of proteins from peptides

MASCOT .dat files containing results from MASCOT searches were uploaded to DBParser [15], a program that compares the MASCOT peptide ions score [ $-10 \cdot \log_{10}(P)$ ] to the MASCOT database-dependent identity score (Iscore), whereby  $P < 0.05$  is the significance threshold for the probability of a false-positive match. DBParser selected for ions scores that met or exceeded the corresponding Iscore, meaning that there was a 5% (5e-2) or lower probability that the observed peptide match was a random event. Subsequently, a protein assignment was accepted if it contained at least one significant assigned peptide and only significant peptides were collated. The program sorted the MASCOT protein assemblies into a NR protein complement by applying a parsimony principle. Data were processed on a 3.2-

GHz Dell desktop computer. Complete DBParser output datasets containing peptide sequences, associated protein information, and their confidence indicators are available as Supplementary Material to this article.

### 3 Results

Our goal was to identify proteins in *U. appendiculatus* uredospores using MS/MS. We isolated loose *U. appendiculatus* uredospores emerging from uredia on infected *P. vulgaris* leaves and then pulverized spores under liquid nitrogen in a mortar and pestle to expose as many internal proteins as possible. Five separate peptide preparations were examined by MudPIT. The .dta files for all five MudPIT experiments were merged into a single file containing 54 535 tandem mass spectra, which was then analyzed by MASCOT against the NCBI NR-protein database. The NCBI NR database was chosen to maximize peptide identification. Based on virtual tryptic peptide sequences derived from the NCBI NR-protein database, MASCOT provided peptide sequence inferences for the observed spectra along with the probability-based scores for match reliability and listed all of the possible-detected proteins with their corresponding inferred peptides in a .dat output file.

MASCOT identified 9188 proteins containing at least one peptide whose virtual spectra matched an *U. appendiculatus* spectrum. However, many of the inferred peptide sequences mapped to more than one protein because NCBI NR contains duplicated protein sequences and protein orthologs, paralogs, or homologs that differ by only a few amino acids. When assessing proteins identified by peptides that map to more than one protein record, hereby defined as degenerate peptides, principles of parsimony dictate that these proteins should be reported as a group until experimental circumstances preclude one protein's existence over another. The identification of unique peptide sequences, hereby defined as discrete peptides that map to only one database record, resolves identification ambiguity. Thus, to analyze the resulting MASCOT output and assemble a minimum set of proteins accounted by the observed proteins, in which proteins identified by the same degenerate peptides are grouped and counted as one protein, we employed the DBParser program. DBParser first determined that the inferred peptides were either discrete or degenerate. Then, DBParser sorted the assembled protein set into a parsimonious NR dataset containing distinct proteins (containing only discrete peptides), differentiable proteins (containing both discrete and degenerate peptides), equivalent proteins (proteins sharing the same set of degenerate peptides), and subsumable proteins (degenerate peptides that can be distributed as subsets of two or more other protein identifications). DBParser reduced the set of 9188 possible redundant proteins to 468 NR proteins (Suppl. Material). There were 190 distinct proteins identified only by discrete peptides; 56 differentiable proteins that had at least one discrete

peptide; 188 NR equivalent proteins that were based on the same set of degenerate peptides (reduced from a total set of 1812 equivalent proteins); and 34 subsumable proteins (Suppl. Material). NR proteins matching Lys-C and trypsin along with obvious nonfungal proteins (contaminants) such as human keratin were manually discarded from the dataset, leaving 462 NR proteins (Suppl. Table 1).

An example of an assembled NR protein is shown in Fig. 1. MASCOT determined that four *U. appendiculatus* spectra most closely matched four virtual spectra from protein UM03791.1 (*Ustilago maydis* 521, gi|46099593), a protein with similarity to other proteins in the HSP70 protein family (Fig. 1a). The false-positive rate ranged from  $1.8 \times 10^{-2}$  to  $1.1 \times 10^{-5}$  for these inferences. One of the peptide sequences inferred from the analyses was present in the NCBI NR database only once, thus making the peptide discrete. By virtue of there being three other degenerate peptides, the assembled protein was classified as a differentiable protein. Together, the four-peptide inferences support our conclusion that a protein similar to UM03791.1 is present in *U. appendiculatus* uredospores.

An analysis of two other NR proteins assemblies is shown in Fig. 2. MASCOT determined that two *U. appendiculatus* spectra most closely matched two virtual spectra from two peptides from a translation elongation factor 2 (EF2) protein from *Neurospora crassa* (gi|13925370). MASCOT also matched the same spectra to the same peptides found in another *N. crassa* EF2 protein (gi|32415856) also present in the database. Although these *N. crassa* EF2 proteins share 838/844 amino acid identity, these proteins cannot be distinguished from each other based solely on these two peptide sequences. Therefore, DBParser grouped the proteins and counted them together as one equivalent protein finding (DBParser arbitrarily assigned gi|13925370 as the representative of this equivalent protein group). MASCOT also determined that one of the *U. appendiculatus* spectra that matched virtual spectra found in the *N. crassa* EF2 proteins could also be found in gi|50258869, an EF2 protein from *Cryptococcus neoformans* var. *neoformans* B-3501A. However, a separate spectrum from *U. appendiculatus* was matched to a different peptide from the *C. neoformans* EF2 protein and this peptide sequence was sufficient to distinguish the *C. neoformans* EF2 protein from the *N. crassa* EF2 proteins. Thus, DBParser also counted the *C. neoformans* EF2 protein separately in the NR set. Our bioinformatics analysis suggests that there are at least two similar EF2 proteins in *U. appendiculatus*. However, further experimental analysis will be required to prove whether there is one or several homologous *U. appendiculatus* EF2 proteins that harbors the three peptide sequences that are also found in the EF2 proteins in *C. neoformans* and *N. crassa*.

Only three protein assemblies were made from known *Uromyces* spp. proteins, which is not surprising given that there are only 19 *Uromyces* spp. protein sequences in the 01/19/2005 release of NCBI NR. The proteins detected were succinate dehydrogenase,  $\beta$ -tubulin, and PIG8, a probable



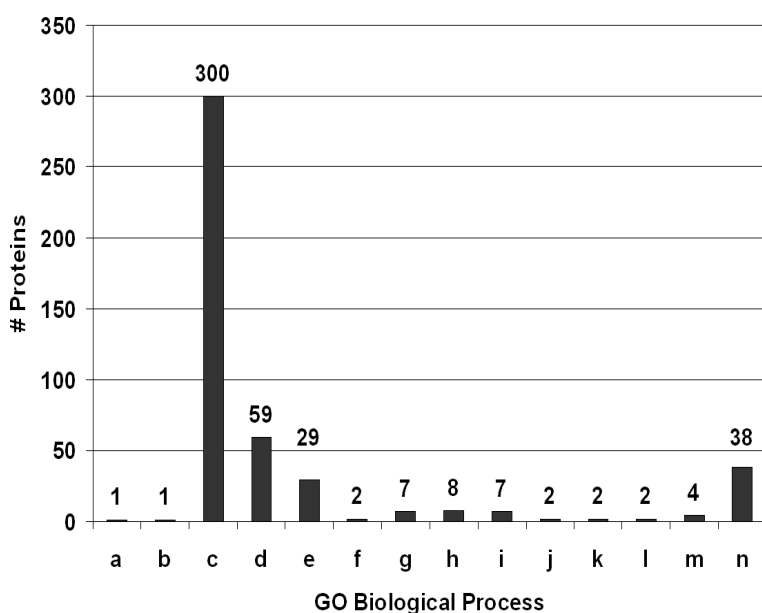
NADP-dependent mannitol dehydrogenase whose gene was originally isolated from a haustoria-specific library [7]. This finding suggests that PIG8 accumulation is not limited to haustoria and, in general, portends the ability of our procedure to localize proteins to cellular structures. Conversely, some of the other *U. appendiculatus* proteins reported in NCBI NR were not specifically detected. It may be possible that further sampling of uredospore protein extracts will be required to detect some proteins like INF56 (gi|170560), which is reported to be isolated from uredospores [6]. We also did not detect some of the proteins reported to be involved in appressorium development, such as INF24 (gi|83858) [5]. Either INF24 is truly not present in uredospores or was not resolved by our experiments.

Most of the other assemblies were derived from proteins from a variety of taxonomically unrelated prokaryote and eukaryote species, although some assemblies were from proteins of other fungi such as *Aspergillus nidulans* (11), *N. crassa* (9), *Yarrowia nidulans* (7), *Magnaporthe grisea* (3), and the phylum-related Basidiomycetes *Us. maydis* (25) and *C. neoformans* (14). These results suggest that there are many peptide sequences that are common between proteins in *U. appendiculatus* and these other organisms.

One concern of ours was assessing our confidence in the veracity of any given protein assembly. We worked with the hypothesis that the assignment of multiple peptides to a protein gives greater confidence for the reliability of a match [24, 25]. Conversely, if only a single peptide is detected, then there is less confidence by comparison, unless that peptide is assigned to only one database record (discrete peptide) at which point it can be considered a unique identifier for a protein [15, 26]. Two hundred and twenty-five of the NR proteins assembled had at least one discrete peptide (Suppl. Material). Of the remaining proteins without discrete pep-

tides, 92 had multiple peptides associated with them. Thus, 317 proteins were assembled based on peptide inferences from unique sequences or from inferences of multiple peptides spanning the protein. We believe that the proteins found in this set are likely to exhibit relatedness to proteins in *U. appendiculatus*.

To gain a general view of the function of the many proteins that were assembled, we categorized them by their Gene Ontology (GO) Biological Process descriptions (<http://www.geneontology.org/>; [27]), which describe an ordered assembly of molecular functions. The Biological Process groupings (Fig. 3, Suppl. Table 1) suggest that there are many proteins in uredospores that are important for protein production, represented by proteins with functions for amino acid metabolism (phosphoribosyl anthranilate isomerase and aspartate transaminase), protein biosynthesis (ribosomal subunits 40S, 60S, L3, and L27), protein translational initiation (EIF-4A), protein translational elongation (elongation factor 1-alpha), protein folding (HSP70, HSP90, dnaK, immunoglobulin heavy chain binding protein), and ubiquitin-dependent protein catabolism (26S proteasome regulatory subunit and polyubiquitin). Seventy-seven different heat shock proteins and molecular chaperones, 27 different ribosomal subunit proteins, and 28 different translation elongation factors appeared in the dataset. Processes for growth and development were represented by proteins that function in cell cycling (cdc48, EB1), biogenesis (actin), gene transcription (myb transcription factor), and proton transport/unidimensional cell growth (ATP synthase beta subunit and V-type ATPase). Twenty-nine ATPase subunit proteins, 19 actin proteins, 10  $\beta$ -tubulin proteins, and 9 transcription factors/regulators were part of the dataset. Other processes, which may be controlled by the proteins identified, include chemotaxis, ergosterol biosynthesis, lipo- and exopoly-



**Figure 3.** Distribution of inferred proteins as grouped by their associated GO Biological Process subcategorization. (a) cellular process/cell communication/signal transduction; (b) cellular process/cellular physiological process/cell death; (c) cellular process/cellular physiological process/cellular metabolism; (d) cellular process/cellular physiological process/cellular organization and biogenesis; (e) cellular process/cellular physiological process/regulation of cellular physiological process; (f) cellular process/physiological process; (g) cellular process/physiological process/cell cycle; (h) cellular process/physiological process/transport; (i) physiological process/metabolism/macromolecule metabolism; (j) physiological process/response to stimulus/response to abiotic stimulus; (k) physiological process/response to stimulus/response to stress; (l) regulation of biological process/positive regulation of biological process/positive regulation of enzyme activity; (m) regulation of biological process/regulation of enzyme activity/regulation of oxidoreductase activity; (n) unknown.

saccharide synthesis, nonribosomal peptide synthesis, peptidoglycan biosynthesis, and membrane transport. Proteins in these latter classes may be important for forming or mobilizing fungal-specific compounds needed for plant infection.

## 4 Discussion

We have used MudPIT to identify the types of proteins that are in uredospores of *U. appendiculatus*, a fungus for which very little genome information exists. The proteins that we present are based on the identification of peptides from homologous proteins from other organisms. Protein homology has already been used successfully to interpret spectra from other organisms for which little genome information is available. For example, a rice protein database was used for the identification of nearly half of 141 proteins of the wheat amyloplast detected by MS/MS [22], and the NCBI NR-protein database was used for the identification of tandem mass spectra from the unsequenced *Candida magnoliae* [28]. Only three of our assemblies were made from *Uromyces* spp. proteins, meaning that nearly all of our protein identifications were based on spectral inferences to homologous peptides found in proteins of different organisms. For example, 25 of our assemblies were made from proteins from *Us. maydis*, another plant pathogenic fungus. Because the *Us. maydis* genome is sequenced, its proteome was well represented in NCBI NR compared to *U. appendiculatus*. It is also a related Basidiomycetes fungus. Therefore, the extensive genome information for *Us. maydis* and the (likely) homology between the proteomes of the two organisms enabled the inferences of the *U. appendiculatus* protein tandem mass spectra. The same can be said for the other fungal organisms whose proteomes are also well defined. The assemblies made to proteins from organisms that are not fungi may also be indicative of the evolutionary divergence of some of the other *U. appendiculatus* proteins.

We performed our searches against the NCBI NR-protein database to maximize our potential for identifying a wide variety of possible peptides that could be in uredospores. One drawback of searching against such a large database is the generation of false-positive identifications [29]. Several spectral interpretation programs, including MASCOT used in this study, apply statistical standards and probability-based models to ascertain the quality of matches and increase the confidence of obtaining quality inferences [30–33]. Similar approaches have been taken for the assembly of proteins from the peptide inferences [24]. We have used both MASCOT and DBParser to collate peptide inferences with database-dependent scores indicating the probability of a random match. These actions help reduce false inferences [34]. Our ability to define distinct proteins assembled from discrete peptides

assigned to only one record also gives greater assignment confidence [35]. In addition, we have combined the collected spectra from redundant runs for data analysis, which provides increased information for assemblies and is proven to result in a better representation of proteins available in the samples [35–37].

We have included in our report single peptide protein assignments of equivalent and subsumable proteins whose sequences appear in more than one protein record (degenerate peptides). While the possibility of assigning single peptides to proteins can be associated with an increased incidence of false identifications [25, 29], it may not be necessary to disregard proteins identified by single peptides outright, especially if the scope of a global proteomics experiment is set for survey or discovery [38]. For our purposes, single degenerate peptide inferences may be useful for identifying equivalent or subsumable proteins since degenerate peptides can indicate regions of protein homology [39].

Our results suggest that there is a variety of protein homologs in the uredospores, *i.e.*, there are matches to 77 different heat shock related proteins, matches to 19 separate actin homologs and 10  $\beta$ -tubulin homologs from other organisms. In *Arabidopsis thaliana*, a plant whose genome has been sequenced, there are more than 180 different heat shock related and approximately 24 different actin proteins. In *Us. maydis*, there are approximately eight different actin-like proteins and four different  $\beta$ -tubulin-like proteins. Thus, finding many different types of heat shock proteins or actin homologs in uredospores is not surprising. However, there is a possibility that the numbers for each class may be artificially inflated as result of searching a large database such as NCBI NR rather than a specific database composed of *U. appendiculatus* proteins. Had we been able to restrict our MASCOT searches to a *U. appendiculatus* comprehensive protein database, then we would have likely obtained a more realistic number of protein homologs detected. Nonetheless, because MudPIT is biased toward detecting abundant proteins and there is a direct correlation between spectral count, peptide count, and sequence coverage [36], we concomitantly believe that we have detected an abundance of heat shock proteins and translation elongation factors in uredospores because many different spectra have been assigned to peptides from these proteins.

Identification of specific classes of proteins in this study may have some relation to the biology of the uredospores. When a *U. appendiculatus* uredospore germinates in water, a germ tube grows, into which moves the cytoplasm and two nuclei [40]. After the appressorium forms as a result of thigmotropic response [4], the cytoplasm and nuclei move into it and mitosis occurs, and septum formation is completed to separate the appressorium from the empty germ tube [41]. Then an infection peg penetrates the stomata, which subsequently forms a substomatal vesicle where a second round of

mitosis can occur [2, 40]. This growth and the two rounds of nuclear replication occur before the formation of an infection hyphae, a haustorial mother cell and haustoria, the organ that absorbs and uptakes nutrients from the plant cell [8]. Thus, it is presumed that the germling does not obtain nutrients from the host until formation of haustoria, and that all of the energy and resources needed to accomplish the events leading up to this point must be initially contained in the uredospores. This means that uredospores must produce proteins very quickly to accommodate all of these processes. We detected an abundance of translation elongation factors in uredospores in addition to a complement of other proteins required for protein production, such as translation initiation and termination factors, ribosomal proteins, amino acid synthases, amino acid modifying enzymes, tRNA synthetases, PTM enzymes, and protein degradation components. Finding a preponderance of proteins related to protein catabolism agrees with the supposition that uredospores need to produce proteins very rapidly upon germination and are prepared to do so.

This notion is supported by finding a large number of heat shock proteins and chaperones in the uredospores. In addition to regulating heat shock responses [42, 43], assisting in transporting proteins across membranes [44], and catalyzing the disassembly of clathrin-coated vesicles [45], 70-kDa heat shock proteins co-operate with translation elongation factors and ribosomal proteins to maintain the integrity of the folding confirmation of nascent polypeptides before they are released from the ribosome [46–49]. Thus, we propose a model by which a uredospore exists in a suspended translational state. We hypothesize that upon germination, uredospores are able to produce proteins very quickly without having to assemble translational components *de novo*. Likewise, peptide chains, suspended in the ribosome, are protected by heat shock proteins and better able to resist adverse conformation changes caused by desiccation or high temperatures to which uredospores might be subjected once released into the environment.

We anticipate that these results will further our understanding of rust biology and expect to use similar approaches in elucidating proteins that comprise different cellular structures that allow the fungus to colonize a plant, absorb nutrients, or reproduce. Future studies will also include obtaining more information about the proteins that are undefined in function, 38 of which were inferred in this study. The *U. appendiculatus* homologs to these proteins certainly are candidates for belonging to a group of species-specific proteins that are responsible for pathogenicity or host recognition.

We thank Dr. Talo Pastor-Corrales, USDA-ARS, Beltsville, for *U. appendiculatus* spores, Dr. Xiaoyu Yang, NIH, for DBParser technical assistance, Chris Overall for SGMD website assistance and Jian Feng and Drs. Paul Haynes, Neerav Padliya and Joohyun Lee for help in preparation of the manuscript.

## 5 References

- [1] Agrios, G. N., *Plant Pathology*, Academic Press, San Diego 2004.
- [2] Staples, R. C., Hoch, H. C., in: Hedin, P. A., Menn, J. J., Hollingworth, R. M. (Eds.), *Biotechnology for Crop Protection*, American Chemical Society, Washington, D.C. 1988, pp. 82–93.
- [3] Bonde, M. R., Melching, J. S., Bromfield, K. R., *Phytopathology* 1976, **66**, 1290–1294.
- [4] Hoch, H., Staples, R., Whitehead, B., Comeau, J. *et al.*, *Science* 1987, **235**, 1659–1662.
- [5] Bhairi, S. M., Staples, R. C., Freve, P., Yoder, O. C., *Gene* 1989, **81**, 237–243.
- [6] Xuei, X., Bhairi, S., Staples, R. C., Yoder, O. C., *Gene* 1992, **110**, 49–55.
- [7] Hahn, M., Mendgen, K., *Mol. Plant Microbe. Interact.* 1997, **10**, 427–437.
- [8] Voegelé, R. T., Struck, C., Hahn, M., Mendgen, K., *Proc. Natl. Acad. Sci. USA* 2001, **98**, 8133–8138.
- [9] Schillberg, S., Tiburzy, R., Fischer, R., *Mol. Gen. Genet* 2000, **262**, 911–915.
- [10] Barja, F., Correa, A. J., Staples, R. C., Hoch, H. C., *Mycol. Res.* 1998, **102**, 1513–1518.
- [11] Florens, L., Washburn, M. P., Raine, J. D., Anthony, R. M. *et al.*, *Nature* 2002, **419**, 520–526.
- [12] Hall, N., Karras, M., Raine, J. D., Carlton, J. M. *et al.*, *Science* 2005, **307**, 82–86.
- [13] Washburn, M. P., Wolters, D., Yates, J. R. III, *Nat. Biotechnol.* 2001, **19**, 242–247.
- [14] Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S., *Electrophoresis* 1999, **20**, 3551–3567.
- [15] Yang, X., Dondeti, V., Dezube, R., Maynard, D. M. *et al.*, *J. Proteome Res.* 2004, **3**, 1002–1008.
- [16] Koller, A., Washburn, M. P., Lange, B. M., Andon, N. L. *et al.*, *Proc. Natl. Acad. Sci. USA* 2002, **99**, 11969–11974.
- [17] Durr, E., Yu, J., Krasinska, K. M., Carver, L. A. *et al.*, *Nat. Biotechnol.* 2004, **22**, 985–992.
- [18] Gatlin, C. L., Kleemann, G. R., Hays, L. G., Link, A. J. *et al.*, *Anal. Biochem.* 1998, **263**, 93–101.
- [19] Yates, J. R. III, McCormack, A. L., Link, A. J., Schieltz, D. *et al.*, *Analyst* 1996, **121**, 65R–76R.
- [20] Wu, C. C., MacCoss, M. J., Howell, K. E., Yates, J. R. III, *Nat. Biotechnol.* 2003, **21**, 532–538.
- [21] Link, A. J., Eng, J., Schieltz, D. M., Carmack, E. *et al.*, *Nat. Biotechnol.* 1999, **17**, 676–682.
- [22] Andon, N. L., Hollingworth, S., Koller, A., Greenland, A. J. *et al.*, *Proteomics* 2002, **2**, 1156–1168.
- [23] Haynes, P. A., Fripp, N., Aebersold, R., *Electrophoresis* 1998, **19**, 939–945.
- [24] Nesvizhskii, A. I., Keller, A., Kolker, E., Aebersold, R., *Anal. Chem.* 2003, **75**, 4646–4658.
- [25] Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J. *et al.*, *J. Proteome Res.* 2003, **2**, 43–50.
- [26] Nesvizhskii, A. I., Aebersold, R., *Mol. Cell Proteomics* 2005, **4**, 1419–1440.
- [27] Clark, J. I., Brooksbank, C., Lomax, J., *Plant Physiol.* 2005, **138**, 1268–1279.

- [28] Kim, H. J., Lee, D. Y., Lee, D. H., Park, Y. C. *et al.*, *Proteomics* 2004, 4, 3588–3599.
- [29] Cargile, B. J., Bundy, J. L., Stephenson, J. L., Jr., *J. Proteome Res.* 2004, 3, 1082–1085.
- [30] Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L. *et al.*, *J. Proteome Res.* 2004, 3, 958–964.
- [31] Sadygov, R. G., Liu, H., Yates, J. R., *Anal. Chem.* 2004, 76, 1664–1671.
- [32] Keller, A., Nesvizhskii, A. I., Kolker, E., Aebersold, R., *Anal. Chem.* 2002, 74, 5383–5392.
- [33] Tabb, D. L., Narasimhan, C., Strader, M. B., Hettich, R. L., *Anal. Chem.* 2005, 77, 2464–2474.
- [34] Nesvizhskii, A. I., Aebersold, R., *Drug Discov. Today* 2004, 9, 173–181.
- [35] Maynard, D. M., Masuda, J., Yang, X., Kowalak, J. A. *et al.*, *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 2004, 810, 69–76.
- [36] Liu, H., Sadygov, R. G., Yates, J. R. III, *Anal. Chem.* 2004, 76, 4193–4201.
- [37] Brechi, L., Hattrup, E., Keeler, M., Letarte, J. *et al.*, *Proteomics* 2005, 5, 2018–2028.
- [38] Veenstra, T. D., Conrads, T. P., Issaq, H. J., *Electrophoresis* 2004, 25, 1278–1279.
- [39] Rappsilber, J., Mann, M., *Trends Biochem. Sci.* 2002, 27, 74–78.
- [40] Heath, I. B., Heath, M. C., *J. Cell Biol.* 1976, 70, 592–607.
- [41] Kwon, Y. H., Hoch, H. C., *Exp. Mycol.* 1991, 15, 116–131.
- [42] Abravaya, K., Myers, M. P., Murphy, S. P., Morimoto, R. I., *Genes Dev.* 1992, 6, 1153–1164.
- [43] Schroder, H., Langer, T., Hartl, F. U., Bukau, B., *Embo. J.* 1993, 12, 4137–4144.
- [44] Herrmann, J. M., Neupert, W., *Curr. Opin. Microbiol.* 2000, 3, 210–214.
- [45] Schmid, S. L., Braell, W. A., Rothman, J. E., *J. Biol. Chem.* 1985, 260, 10057–10062.
- [46] Beckmann, R. P., Mizzen, L. E., Welch, W. J., *Science* 1990, 248, 850–854.
- [47] Frydman, J., Nimmesgern, E., Ohtsuka, K., Hartl, F. U., *Nature* 1994, 370, 111–117.
- [48] Nelson, R. J., Ziegelhoffer, T., Nicolet, C., Werner-Washburne, M. *et al.*, *Cell* 1992, 71, 97–105.
- [49] Wegele, H., Muller, L., Buchner, J., *Rev. Physiol. Biochem. Pharmacol.* 2004, 151, 1–44.